

## Special Session on Large-Scale Clustering

[2011 IEEE Conference on Fuzzy Systems](#), Taipei, Taiwan, June 27-30, 2011

Since the early 1990s, the ubiquity of personal computing technology has produced an abundance of staggeringly large data sets – the Library of Congress has stored over 160 terabytes of web data and it is estimated that Facebook alone logs over 25 terabytes of data per day. There is a great need for systems by which one can elucidate the similarity and dissimilarity among and between groups in these data sets and produce easy-to-understand visualizations or interpretations of the results. One such method that can be used to attack large-scale data problems is clustering.

The aim of this special session is to present leading research pertaining to the theory and application of clustering with large-scale data sets. This session will also provide a forum by which the academy and industry can report on recent advances in the field.

Topics for submissions can include, but are not limited to, the following:

- Scalable clustering / partitioning algorithms
- Cluster tendency — How many clusters are there?
- Cluster validity measures
- Cluster visualization methods
- Interpretation of clustering results — knowledge discovery
- Partition comparison — Which partition is better?
- Choosing a fuzzy, crisp, probabilistic, possibilistic, or mixed-membership clustering algorithm — Is there a “best” algorithm for a given data set?
- Clustering as a pre-processing method for other large-scale algorithms, such as classification or indexing
- Performance evaluation of clustering algorithms
- Scalable kernel clustering
- Ensemble methods for large-scale clustering
- Applications of clustering in large-scale data

Session organizer:

Dr. Timothy Havens

Department of Computer Science and Engineering

Michigan State University

3115 Engineering Building

East Lansing, MI 48824

Phone: 231-360-8444

E-mail: [havenst@msu.edu](mailto:havenst@msu.edu)